



**HAL**  
open science

## Diagnostic automatisé d'aiguillage ferroviaire par apprentissage statistique

Quentin Grimonprez, Benjamin Lamoureux, Etienne Goffinet, Pierre Dersin,  
Christophe Biernacki

### ► To cite this version:

Quentin Grimonprez, Benjamin Lamoureux, Etienne Goffinet, Pierre Dersin, Christophe Biernacki.  
Diagnostic automatisé d'aiguillage ferroviaire par apprentissage statistique. Congrès Lambda Mu 22  
“ Les risques au cœur des transitions ” (e-congrès) - 22e Congrès de Maîtrise des Risques et de Sécurité  
de Fonctionnement, Institut pour la Maîtrise des Risques, Oct 2020, Le Havre (e-congrès), France.  
hal-03479596

**HAL Id: hal-03479596**

**<https://hal.science/hal-03479596>**

Submitted on 14 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Diagnostic automatisé d'aiguillage ferroviaire par apprentissage statistique

## Automated diagnosis of railway switches by statistical learning

Quentin Grimonprez  
Équipe Inria MODAL  
Inria Lille-Nord Europe  
Villeneuve d'Ascq, France  
quentin.grimonprez@inria.fr

Etienne Goffinet  
Équipe Inria MODAL  
Inria Lille-Nord Europe  
Villeneuve d'Ascq, France

Christophe Biernacki  
Équipe Inria MODAL  
Inria Lille-Nord Europe  
Villeneuve d'Ascq, France

Benjamin Lamoureux  
Alstom  
Saint-Ouen, France

Pierre Dersin  
Alstom  
Saint-Ouen, France

**Résumé**—Le but de cette étude est de comparer les performances des méthodes d'analyse de données fonctionnelles par rapport à une approche basée sur des descripteurs métiers pour automatiser le diagnostic d'aiguillages ferroviaires.

**Mots-clés**—*diagnostic, classification, données fonctionnelles, analyse en composantes principales*

**Abstract**—The purpose of this study is to compare the performance of functional data analysis methods against a descriptor-based approach to automate the diagnosis of railway switches.

**Keywords**—*prognostic, clustering, functional data, principal component analysis*

### I. INTRODUCTION

La surveillance de l'état des équipements répartis sur le réseau ferré, tels les aiguillages, est assurée au moyen de tournées d'inspection de techniciens. Ces tournées d'inspection sur le réseau ferroviaire représentent un coût important mais restent indispensables pour garantir la disponibilité du réseau et la sécurité des usagers. Afin de combiner un suivi en continu plus sûr et des tournées d'inspection moins fréquentes, une étude sur le diagnostic à distance de l'état de fonctionnement d'aiguillages a été mise en œuvre en commençant par l'acquisition de données sur un banc test.

Le jeu de données moteur d'aiguilles (MDA) étudié dans cet article provient d'enregistrements effectués par ALSTOM sur ce banc test. Ce jeu contient des informations issues de capteurs sur des aiguillages ferroviaires. Les capteurs

enregistrent 3 paramètres de fonctionnement (vitesse, couple, puissance électrique consommée) du moteur déplaçant l'aiguille lorsque celui-ci est actionné. Chaque manœuvre dure 10 secondes avec un enregistrement toutes les millisecondes engendrant ainsi des données fonctionnelles pour chaque paramètre et chaque manœuvre d'aiguillage. Environ 20000 manœuvres ont été enregistrées et annotées avec 4 statuts (dont 3 dysfonctionnements) : sain, frottement, désalignement et obstacle. Des courbes de consommation de diverses manœuvres avec divers états sont visibles sur la figure 1. Le but de l'étude est de diagnostiquer l'état de santé d'une manœuvre à partir de ces données.

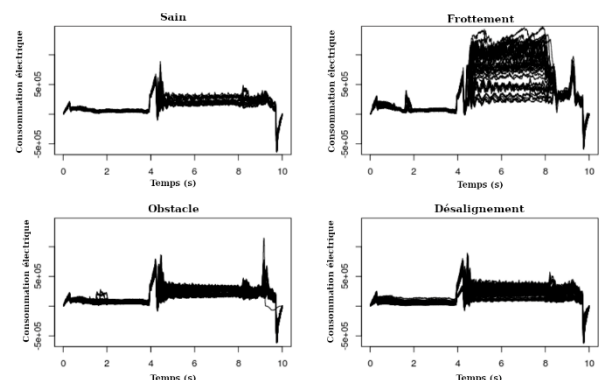


Figure 1: Consommation électrique pour différentes manœuvres. Chaque courbe correspond à une manœuvre.

Dans de précédentes études, chaque courbe était découpée en différentes zones pour lesquelles des descripteurs (moyenne, minimum, maximum...) étaient extraits et utilisés lors d'analyses (cf. figure 2). Cette phase de prétraitement présente quelques inconvénients comme la perte d'information, l'introduction de biais technique ou encore son temps de développement. Les contraintes de cette étude sont donc de traiter directement les données dites fonctionnelles pour prédire le statut tout en utilisant des modèles interprétables.

Le concept de l'analyse de données fonctionnelles est de considérer l'ensemble des enregistrements aux différents temps comme une seule entité et non comme un ensemble d'enregistrements univariés. L'analyse de données fonctionnelles s'est rapidement développée suite au développement des capteurs et de l'informatique. Les bases de ce champ de recherche sont regroupées dans [1]. Nous nous intéressons à diverses méthodes statistiques en présence de données fonctionnelles. Tout d'abord, aux méthodes de classification non supervisée afin de déterminer les profils majoritaires de consommations électriques. Puis à celles de classification supervisée afin d'apprendre un modèle capable de diagnostiquer l'état de santé d'un aiguillage. Différentes approches existent pour la classification non supervisée : la classification ascendante hiérarchique [2], les k-means [3], ou encore les modèles de mélange [4]. Ces derniers sont une approche basée sur une modélisation des données par une loi de probabilité. Nous les présentons dans la partie II. L'autre problème qui nous intéresse est celui de la classification supervisée en présence de variables explicatives fonctionnelles afin de pouvoir prédire l'état de santé des aiguillages en fonction de la consommation électrique. Pour cela, nous nous tournons vers les méthodes de régressions fonctionnelles [5] qui seront présentées dans la partie III. L'ensemble des méthodes présentées sont testées sur le jeu de données MDA dans la partie IV. La partie V conclura l'étude.

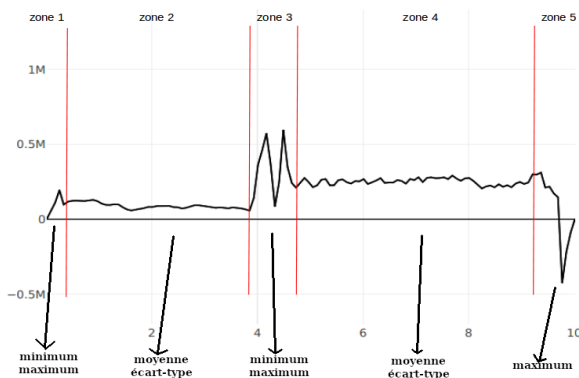


Figure 2: Découpage en zones et descripteurs.

## II. MODELES DE MELANGE POUR LA CLASSIFICATION NON SUPERVISEE DE DONNEES FONCTIONNELLES

### A. Modèles de mélange

Les modèles de mélange sont une approche probabiliste du problème de classification non supervisée. Ils supposent que la population est issue d'une somme pondérée de lois de probabilité. Chacune de ces lois correspondant à une classe de la population. Notons  $x_1, \dots, x_n$ ,  $n$  observations, la loi mélange pour un individu  $x_i$  est définie par :

$$p(x_i; \theta) = \sum_{k=1}^K \pi_k p(x_i; \alpha_k)$$

avec  $K$  le nombre de classes,  $\pi_k$  les proportions du mélange telles que  $0 \leq \pi_k \leq 1$  et  $\sum_{k=1}^K \pi_k = 1$  et  $p(\cdot; \alpha_k)$  la loi de la  $k$ -ième classe paramétrisée par  $\alpha_k$  et  $\theta = (\pi_1, \dots, \pi_K, \alpha_1, \dots, \alpha_K)$ , l'ensemble des paramètres.

Le problème de classification non supervisée se résume alors à estimer les paramètres  $\theta$ . Cette estimation se fait souvent par maximisation de la vraisemblance grâce à un algorithme de type EM [6] ou SEM [7]. Une fois les paramètres estimés, la partition des données est obtenue en calculant la probabilité conditionnelle d'être dans chacune des classes :

$$t_{ik} = \frac{p(x_i; \alpha_k)}{\sum_{j=1}^K \pi_j p(x_i; \alpha_j)}$$

Puis, la classe  $z_i$  de  $x_i$  est déterminée par maximum a posteriori (MAP) :

$$z_i = \operatorname{argmax} t_{ik}.$$

Les modèles de mélanges sont très flexibles et permettent de traiter des données multivariées et hétérogènes (variables de natures différentes : continues, catégorielles, entières, fonctionnelles...) facilement. Notons  $x_i = (x_i^1, \dots, x_i^d)$ , un individu multivarié hétérogène. Alors, sous l'hypothèse d'indépendance conditionnelle aux classes :

$$p(x_i; \alpha_k) = \prod_{j=1}^d p(x_i^j; \alpha_k^j)$$

où  $p(\cdot; \alpha_k^j)$  est la loi associée à la  $j$ -ième variable,  $\alpha_k^j$  les paramètres associés à cette loi et  $\alpha_k = (\alpha_k^1, \dots, \alpha_k^d)$  l'ensemble des paramètres à estimer, on a :

$$p(x_i; \theta) = \sum_{k=1}^K \pi_k p(x_i; \alpha_k) = \sum_{k=1}^K \pi_k \prod_{j=1}^d p(x_i^j; \alpha_k^j).$$

Sous cette hypothèse, il suffit donc de disposer d'une loi univariée pour chaque type de variable. Ainsi pour faire de la classification non supervisée de données fonctionnelles multivariées, une loi pour données fonctionnelles doit être définie.

### B. Distribution pour données fonctionnelles

Diverses approches pour classer les données fonctionnelles via les modèles de mélanges existent [8]; nous proposons d'utiliser la loi présentée dans [9]. Les auteurs proposent d'utiliser un modèle de mélange de régression pour modéliser les données fonctionnelles. Cela consiste à décomposer une courbe en plusieurs segments (ou périodes temporelles) à l'aide de polynômes de degré  $p$ . Par exemple, sur la figure 3, une courbe est décomposée en 5 polynômes (un par segment) de degré 3 fournissant ainsi un lissage de la courbe.

La loi est ainsi paramétrée par ce mélange de polynômes jouant le rôle de moyenne et d'un bruit gaussien pour modéliser la variance autour de cette moyenne.

Notons  $y = (y_1, \dots, y_m)$  une courbe de  $m$  observations univariées observées aux temps  $t = (t_1, \dots, t_m)$ . La densité de probabilité est définie par :

$$f(y | t; \theta) = \prod_{j=1}^m \sum_{l=1}^L \kappa_l(t_j; \alpha) N(y_j; \beta_l^T T_j, \sigma_l^2)$$

avec  $L$  le nombre de segments,  $\kappa_l(t_j; \alpha)$  la probabilité que  $t_j$  appartiennent au segment  $l$ ,  $T_j$  le vecteur tel que  $T_j = (1, t_j, \dots, t_j^p)^T$  avec  $p$  le degré polynomial,  $N(\cdot; \mu, \sigma^2)$  la densité d'une loi gaussienne de moyenne  $\mu$  et variance  $\sigma^2$ , et  $\theta = \{\alpha, \beta_1, \dots, \beta_L, \sigma_1, \dots, \sigma_L\}$  les paramètres de la distribution à estimer,  $p$  et  $L$  étant des hyperparamètres fixés par l'utilisateur.

Un des avantages de cette loi est d'être facilement interprétable puisque que les paramètres permettent de fournir une courbe moyenne et une variance sur chaque segment.

Ce modèle est implémenté dans le logiciel RMixtComp [10] que nous utilisons dans la partie IV.

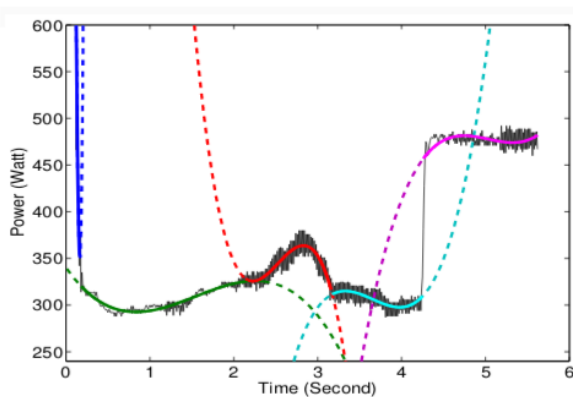


Figure 3: Courbe décomposée en 5 polynômes de degré 3. Image issue de [9].

### III. MODELES PREDICTIFS POUR DONNEES FONCTIONNELLES

Pour appliquer une régression multinomiale avec une variable fonctionnelle comme variable explicative, nous ramenons le problème à celui d'une régression multinomiale avec des variables réelles par le biais de la projection des données dans une base de fonctions et de l'analyse en composantes principales sur données fonctionnelles (ACPf) que nous présentons dans la suite.

#### A. Projection dans une base de fonctions

En analyse de données fonctionnelles, on suppose que les enregistrements aux différents temps sont les observations d'une fonction. On dispose ainsi de couples  $(t_j^i, y_j^i)$  pour  $j = 1, \dots, p$  qui sont des observations d'une fonction  $f_i$ . Il faut donc dans un premier temps reconstruire cette fonction à partir des données. Une des approches est d'utiliser une base de fonctions afin de projeter les données.

L'idée de la projection est de supposer que la fonction  $f_i$  appartient à un espace généré par un ensemble de  $B$  fonctions appelées fonctions de bases  $\{\varphi_b, b = 1, \dots, B\}$ , avec  $B \geq 1$  le nombre de fonctions de bases. Ainsi on représente  $f_i$  pour un temps  $s, t_1 \leq s \leq t_m$  :

$$f_i(s) = \sum_{b=1}^B \alpha_b^i \varphi_b(s)$$

avec  $\alpha_b^i, b=1, \dots, B$  des coefficients réels. Deux bases sont principalement utilisées : la base de Fourier pour représenter les fonctions périodiques et la base B-splines pour les non-périodiques. Sur la figure 4, deux courbes de consommation électrique et leur projection dans une base de B-splines de 101 fonctions de base sont visibles. On note ici, un effet de lissage obtenu après projection. Plus le nombre de fonctions de base utilisé est petit, plus cet effet de lissage est important.

Cette projection dans une base fournit alors une représentation de la fonction par un ensemble de données réelles : les coefficients  $\alpha_b^i$ . Nous allons modifier cette représentation à l'aide d'une ACP afin d'obtenir une nouvelle représentation avec de meilleures propriétés.

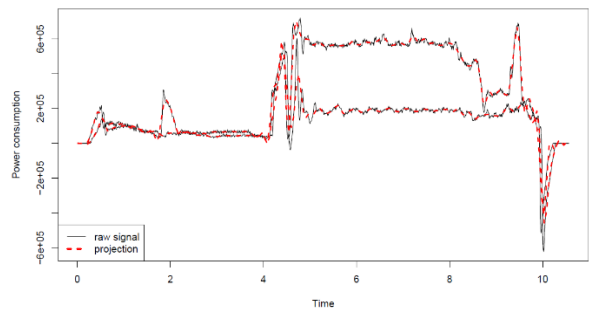


Figure 4: Courbes de consommation électrique (en noir) et leurs projections dans une base B-splines de 101 fonctions de base (en rouge).

## B. Analyse en composantes principales pour données fonctionnelles

L'idée de l'ACP est de résumer les données par un ensemble de composantes principales qui sont une combinaison linéaire des variables, ici  $\{f_i(t), t_1 \leq t \leq t_p\}$ . Dans le cas des données fonctionnelles, cette combinaison s'écrit :

$$c_b^i = \int_{t_1}^{t_p} f_i(t) u_b(t) dt$$

avec les  $u_b$  formant une base orthonormée.

Les  $u_b$  sont les solutions de l'équation

$$\int_{t_1}^{t_p} \text{Cov}(f_i(t), f_i(s)) u_b(s) ds = \lambda_b u_b(t)$$

avec Cov l'opérateur de covariance et  $\lambda_b$  une valeur propre.

Dans le cas où les fonctions sont projetées dans une base de fonctions, il est montré que l'ACP se résume à une ACP sur les coefficients de bases [1].

En utilisant la projection de  $f_i$  dans la base  $\{\varphi_b, b = 1, \dots, B\}$ , l'ACP est équivalente à une ACP sur les coefficients  $\{\alpha_b^i, b = 1, \dots, B\}$  multipliés par une matrice contenant les produits scalaires des fonctions de base [1]. Ainsi, on obtient une représentation de  $f_i$  dans la base  $\{u_b, b = 1, \dots, B\}$  qui est orthonormée :

$$f_i(t) = \sum_{b=1}^B c_b^i u_b(t).$$

Les avantages de cette représentation sont de disposer d'une base orthonormée et de composantes non corrélées qui expliquent au mieux la variance des données. Ainsi, un gain d'interprétabilité en résulte. Par exemple, sur la figure 5 est représentée l'évolution de la courbe moyenne en fonction de la valeur de la première composante de l'ACP sur les données MDA. On constate que la première composante (qui représente une large partie de la variance des données) modifie principalement la valeur moyenne de la fonction entre  $t=5s$  et  $t=9s$ .

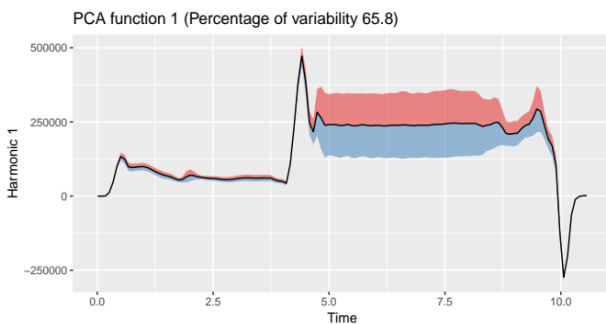


Figure 5: Impact de l'évolution de la valeur de première composante sur la courbe moyenne (en noire). La zone en rouge (resp. bleue) correspond à une valeur positive (resp. négative) égale à 2 fois l'écart-type de la composante.

Un autre avantage de l'ACP est de facilement réduire la dimension en gardant uniquement les composantes expliquant un certain pourcentage de la variance des données.

Cette nouvelle représentation des fonctions  $f_i$  va permettre d'appliquer facilement la régression multinomiale pour des variables fonctionnelles explicatives.

## C. Régression multinomiale avec une variable explicative fonctionnelle

Tout d'abord, rappelons le modèle de régression multinomiale dans le cas de variables réelles. On dispose d'une variable réponse catégorielle  $y = (y_1, \dots, y_n)$  à valeurs dans  $\{1, 2, \dots, J\}^n$  et d'une matrice  $X$  de taille  $n \times p$  contenant les valeurs de  $p$  variables explicatives pour  $n$  individus. La régression multinomiale consiste à apprendre  $J-1$  modèles binomiaux où une catégorie est utilisée comme référence (la catégorie 1 dans notre cas) :

$$\ln \left( \frac{P(y_i = j|X)}{P(y_i = 1|X)} \right) = \beta_0^j + X\beta^j$$

avec  $\beta_0^j \in \mathbb{R}$ ,  $\beta^j \in \mathbb{R}^p$  les coefficients à estimer.

Ces  $J-1$  modèles permettent d'obtenir les probabilités suivantes qui permettent de prédire la catégorie la plus probable pour un individu :

$$P(y_i = 1|X_i) = \frac{1}{\sum_{k=2}^J \exp(\beta_0^k + X_i \beta^k)}$$

et pour  $j = 2, \dots, J$  :

$$P(y_i = j|X_i) = \frac{\exp(\beta_0^j + X_i \beta^j)}{\sum_{k=2}^J \exp(\beta_0^k + X_i \beta^k)}$$

La version fonctionnelle de ce modèle s'obtient en remplaçant le produit scalaire dans un espace euclidien par le produit scalaire dans l'espace des fonctions  $L^2$  :

$$\ln \left( \frac{P(y_i = j|f_i)}{P(y_i = 1|f_i)} \right) = \beta_0^j + \int_{t_1}^{t_p} f_i(t) \beta^j(t) dt.$$

Dans le cas où les données fonctionnelles ont été projetées dans une base et qu'une ACP a été réalisée, le modèle devient [5] :

$$\ln \left( \frac{P(y_i = j|f_i)}{P(y_i = 1|f_i)} \right) = \beta_0^j + \sum_{b=1}^B c_b^i \gamma_b^j$$

où  $c_b^i$  est la  $b$ -ième composante de  $f_i$  et avec  $\beta_0^j \in \mathbb{R}$ ,  $\gamma^j \in \mathbb{R}^B$  les coefficients à estimer.

Le modèle de régression multinomiale avec une variable explicative fonctionnelle revient donc à un modèle de régression multinomiale sur les composantes de l'ACP.

#### IV. APPLICATION AU JEU DE DONNEES MDA

Nous appliquons les méthodes présentées ci-dessus au jeu de données MDA. Les objectifs de l'analyse sont :

1. montrer l'efficacité de l'usage des données fonctionnelles par rapport à l'usage de descripteurs ;
2. montrer l'intérêt de la classification non supervisée en amont de l'usage de modèles prédictifs.

Pour cela, nous proposons la stratégie suivante :

1. Prétraiter les données : extraire des descripteurs ou projeter dans une base de fonctions puis réaliser une ACP fonctionnelle ;
2. Faire une classification non supervisée en  $K$  classes des données courbes/descripteurs ;
3. Calculer la partition des données par MAP ;
4. Pour  $k = 1, \dots, K$ : apprendre une régression multinomiale par classe avec sélection stepwise pour prédire l'état de santé en fonction des données courbes/descripteurs.

Les données sont séparées en trois échantillons : apprentissage (60% des données), test (20%) et validation (20%) afin notamment de tester différentes valeurs d'hyperparamètres. La performance des modèles est évaluée via le taux de bonnes prédictions. La performance du meilleur modèle est calculée sur l'échantillon de validation qui n'a servi ni à l'apprentissage ni à l'optimisation des hyperparamètres.

Le but de la classification non supervisée est d'une part descriptive afin de mieux connaître la population composant les données, d'avoir des sous-populations homogènes pour apprendre les modèles de régression et également de flexibiliser les régressions car on obtient ainsi une régression globalement non-linéaire.

##### A. Cas « descripteurs »

Nous disposons d'une vingtaine de descripteurs créés par les ingénieurs d'Alstom basés sur leurs connaissances métiers (cf. figure 2). Sur la figure 6 est représenté le taux de bonnes prédictions (exprimé entre 0 et 1) en fonction du nombre de classes. On observe que ce taux stagne vers un nombre de classes de 30 et que le maximum est atteint pour 45 classes. On constate une nette différence entre le modèle sans classes ( $< 0.7$ ) et celui avec ( $> 0.9$ ) confirmant ainsi l'intérêt de la classification non supervisée en étape préalable. Ce gain de performances s'explique par le fait que la régression logistique génère des frontières de décision linéaires. Réaliser des régressions logistiques par classes permet de casser cette linéarité en obtenant des frontières linéaires par morceaux donc plus à même de séparer efficacement les individus de labels différents. La performance du meilleur modèle avec des descripteurs est de 93.08 % de bonnes prédictions. Pour ce modèle, les performances par statut sont visibles sur la table 1. Le frottement est le statut avec les meilleures performances (99% de bonnes prédictions). On note également que le modèle ne prédit pas de frottement à tort. Cela n'est pas surprenant car en visualisant les courbes par statut, celles avec le statut frottement sont assez différentes des autres (cf. figure 1). Le statut sain est quant à lui celui avec les plus faibles performances (89%).

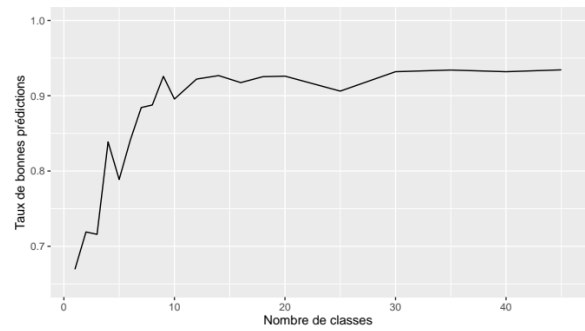


Figure 6: Taux de bonnes prédictions en fonction du nombre de classes dans le cas des descripteurs.

TABLE I. PREDICTION DU STATUT (CAS « DESCRIPTEURS »)

		Prédiction				%
		Frottement	Sain	Obstacle	Désalignement	
V é r i f ié	Frottement	96	0	0	0	99.0
	Sain	0	596	32	42	89.3
	Obstacle	0	26	1591	62	94.8
	Désalignement	0	78	41	1508	92.7

##### B. Cas « fonctionnelles »

Pour le cas des données fonctionnelles, seuls le nombre de classes du modèle de mélange et le nombre de segments de la loi sur données fonctionnelles sont optimisés. L'autre hyperparamètre de la loi, le degré de chaque polynôme, est fixé à 1. Pour la régression multinomiale, la base utilisée est une base B-splines d'ordre 4 (car les signaux ne sont périodiques) avec 101 fonctions de bases. Il est à noter que pour des raisons pratiques (notamment de convergence de l'algorithme), une version hiérarchique de la classification a été faite. On commence par réaliser une classification en 3 classes (profondeur 1) puis une nouvelle classification en 3 classes est réalisée sur les individus de chaque classe séparément et ainsi de suite.

Les taux de bonnes prédictions obtenus pour les différentes valeurs des hyperparamètres sont visibles sur la figure 7. On constate comme pour le cas des descripteurs, que l'utilisation de la classification non supervisée améliore la prédiction passant d'environ 0.88 dans le cas avec une unique classe à environ 0.97 pour le meilleur modèle. Le meilleur modèle est obtenu pour 10 polynômes et une profondeur de 3 avec une performance de 97.30 % de bonnes prédictions. Cette performance est supérieure au cas des descripteurs. En effet, les descripteurs métiers entraînent une perte d'information ne permettant pas de capter certains phénomènes (les descripteurs créés le sont pour détecter des phénomènes connus), ce n'est pas le cas pour les données fonctionnelles. Les performances par statut du modèle sont présentées dans la table 2. Comme pour celui associé aux descripteurs, le statut frottement est celui avec les meilleures performances (99%). Le modèle a des performances similaires pour les 3 autres statuts (~97%).

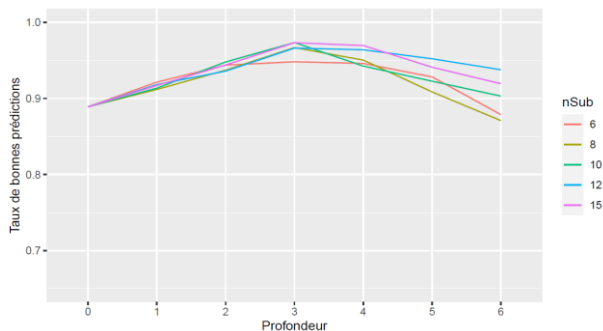


Figure 7: Taux de bonnes prédictions dans le cas des fonctionnelles en fonction de la profondeur et du nombre de segments ( $n_{Sub}$ ).

TABLE II. PREDICTION DU STATUT (CAS « FONCTIONNELLES »)

		Prédiction				%
		Frottement	Sain	Obstacle	Désalignement	
V é r i f i c a t i o n	Frottement	96	0	1	0	99.0
	Sain	0	652	17	1	97.3
	Obstacle	0	21	1641	17	97.7
	Désalignement	0	9	44	1574	96.7

## V. CONCLUSION

La méthode proposée permet de traiter des données de type courbes sans création de descripteurs par les experts métiers. Elle allie classification non supervisée de données courbes par modèle de mélange et régression multinomiale pour données fonctionnelles grâce à la projection dans une base de fonctions et à l'analyse en composantes principales permettant de se ramener à un problème de régression sur données réelles.

Le pouvoir prédictif de la consommation électrique lors d'une manœuvre sur l'état de santé de la manœuvre a ainsi pu être confirmé avec des performances de 97.30 % de bonnes prédictions. L'approche proposée a validé l'intérêt de l'usage des données fonctionnelles à la place de descripteurs par les experts métiers (97.30 % de bonnes prédictions pour les

fonctionnelles contre 93.08 % pour les descripteurs). De plus, l'apport d'une étape de classification non supervisée en amont de l'apprentissage d'un modèle de régression a amélioré les performances en augmentant le taux de bonnes prédictions d'environ 88 % à 97.30 %.

Le jeu de données utilisé dans cette étude provient de bancs tests. Le passage sur données réelles va amener de nouveaux problèmes notamment la diversité des conditions d'utilisations (température, humidité...), le manque d'information sur l'état de santé de l'aiguillage lors de la manœuvre (connu uniquement lorsqu'un technicien passe). Il faudra alors pour chaque aiguillage savoir prédire le passage d'un état sain à un dysfonctionnement. Ce passage pourra se faire de manière progressive via une dérive lente dans le signal (par exemple l'apparition progressive de frottements) ou d'une manière nette (apparition d'un obstacle).

## REMERCIEMENTS

Remerciements à DiagRAMS Technologies, qui exploite les technologies développées par Inria MODAL évoquées dans cette présentation, pour le sponsoring de la participation au congrès  $\lambda\mu 22$ .

## REFERENCES

- [1] J. O. Ramsay and B.W. Silverman, "Functional data analysis," 2nd ed., New York : Springer, ISBN 0-387-40080-X, 2005.
- [2] J Ward, "Hierarchical grouping to optimize an objective function." Journal of the American Statistical Association, 58:236–244, 1963.
- [3] J. Hartigan and M. Wong. "Algorithm as 1326 : A k-means clustering algorithm." Applied Statistics, 28:100–108, 1978,
- [4] J. Banfield, A. Raftery "Model-based gaussian and non-gaussian clustering." Biometrics 49:803–821, 1993.
- [5] J. Morris. "Functional Regression." Annual Review of Statistics and Its Application. 2:321-359, 2014.
- [6] A. Dempster, N. Laird, and D. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society. Series B (Methodological) 39, no. 1 (1977): 1-38.
- [7] G. Celeux, and J. Diebolt. "L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités." Revue de Statistique Appliquée 34.2 (1986): 35-52.
- [8] J. Jacques and C. Preda. "Functional data clustering: a survey". Advances in Data Analysis and Classification, Springer Verlag, 2014, 8 (3), pp.24.
- [9] A. Samé, F. Chamroukhi, G. Govaert and P. Aknin. "Model-Based Clustering and Segmentation of Time Series with Changes in Regime." Advances in Data Analysis and Classification 5.4 (2011): 301–321.
- [10] V. Kubicki, C. Biernacki and Q.Grimonprez. RMixtComp: Mixture Models with Heterogeneous and (Partially) Missing Data. (2020) <https://github.com/modal-inria/MixtComp>